

Proyecto Final MCD412

College Scorecard The Student Debt U.S.

Categoría: Educación
M.Sc. Ing. Leydi Cruz

Julio 2025

Descripción del Dataset

Este dataset contiene información detallada sobre instituciones de educación superior en EE.UU., incluyendo tasas de graduación, deserción, ingresos postgraduación, costos, etc.

El conjunto de datos incluye datos sobre tasas de pago de préstamos estudiantiles, tasas de graduación, asequibilidad, ingresos después de la graduación y más. El objetivo de este conjunto de datos es ayudar a los estudiantes a tomar decisiones informadas sobre su elección universitaria proporcionándoles información clara y concisa sobre el desempeño de cada escuela.

Descripción completa del dataset aquí: [Colleague ScoreCard US Department](#)

Preguntas

¿Qué factores explican mejor la tasa de deserciones universitarias o los ingresos postgraduación en EE.UU.?

¿Es posible predecir cuándo un estudiante va abandonar?

Modelos posibles

Este problema puede ser abordado usando:

- Regresión (ej. ingresos postgraduación)
- Clasificación (ej. alta vs baja deserción)
- Clustering para segmentación institucional

Si bien se sugiere estos modelos, son una guía pero no así una limitante.

Metodología

Para abordar el tema se espera que su análisis tenga como mínimo las siguientes secciones:

1. **Compresión del problema específico.**
2. **EDA:** exploración de correlaciones, normalización, etc. detección de outliers análisis de correlación.
3. **Modelado:**
 - ML básico: regresión lineal / logística.
 - ML avanzado: Random Forest, XGBoost.
 - DL: red neuronal densa.
4. **Evaluación:** Utilizar métricas de evaluación apropiadas por ejemplo RMSE, AUC, precisión, recall, etc.
5. **Reflexión:** El modelo generado es lo suficientemente confiable? Por qué si? Por qué no?Cuál es el potencial impacto del modelo generado?.

Rúbrica de Evaluación

El notebook presentado será evaluado bajo la siguiente rúbrica:

Criterio	Descripción	Puntaje
Planteamiento del problema	Claridad y relevancia	15 pts
EDA	Visualización, insights, limpieza	20 pts
Modelo base ML	Implementación y justificación	15 pts
Modelo avanzado ML	Optimización y mejora	15 pts
Modelo DL	Arquitectura y entrenamiento	15 pts
Evaluación	Métricas adecuadas y análisis	10 pts
Reflexión final	Aplicabilidad, impacto	10 pts
Total		100 pts